

# 基于遗传算法(GA)解甘蔗收割顺序最优化问题的研究\*

孙丽娅<sup>1</sup> 上野正实<sup>2</sup> 大岭政朗<sup>3</sup>

(1. 苏州市职业大学 苏州 215104; 2. 日本琉球大学 日本 冲绳 9010213;  
3. 日本农林水产省九州冲绳农业试验中心 日本 宫崎 8850091)

**摘要** 本研究以已建立的栽培方法、收割时期预测甘蔗单产和含糖量的数学模型为基础,提出将寻求最佳收割顺序、收割期归结为最优化组合问题,并探讨了遗传算法(GA)的解题适用性,同时进行了几种基本设定的栽培方法推移、栽培方法构成项量、行列式特性的定性分析,为求最佳的甘蔗收割顺序、收割期、不同栽培方法的面积等提供解析方法。为不使栽培方法的构成比例发生变化,与通常GA算法采用交叉在不同个体间互相交替进行不同,开发了适用本研究的GA的自己交叉具体解析法。将收割第1年期作为过渡状态,第2年期为安定状态,将两期目的函数的合并作为一个适应度进行评价。

**关键词** 甘蔗 收割顺序 收割期 遗传算法(GA) 自己交叉 最优化

**中图分类号:** S5047; S566.1 **文献标识码:** A **文章编号:** 1671-3990(2009)06-1273-05

## A genetic algorithm methodology for optimized harvesting sequence of sugarcane

SUN Li-Ya<sup>1</sup>, Masami UENO<sup>2</sup>, Masaaki OMINE<sup>3</sup>

(1. Suzhou Vocational University, Suzhou 215104, China; 2. University of the Ryukyus, Okinawa 9010213, Japan; 3. National Agricultural Research for Kyushu Okinawa Region, Miyazaki 8850091, Japan)

**Abstract** New mathematical models capable of predicting sugarcane sugar content based on planting method and harvest time were used to develop a methodology for optimizing harvesting sequence, and harvesting time of sugarcane. The applicability of GA was discussed too. The methodology based on Genetic Algorithm (GA) was used to qualitatively analyze the determinant variables of planting mode to obtain optimized combination harvest sequence, harvest time and area of different planting methods under several setted planting mode transitions. Self-crossing, as opposed to conventional GA crossing methods, was used to retain the structural ratio of planting modes. In this study, first year harvest was used for the transition state and that of the second year for the steady-state runs, while the combination of the two were formed the objective function.

**Key words** Sugarcane, Harvest sequence, Harvest time, Genetic Algorithm, Self-crossing, Optimization

(Received May 8, 2009; accepted July 10, 2009)

不同甘蔗栽培方法与收割时期对甘蔗产量、含糖量及与收割同期进行的春植作业的劳力及机械分配等带来影响。即使是在收获期间,不同栽培方法的甘蔗也以不同程度生长,因而单产及含糖量也在持续变化<sup>[1]</sup>。新植甘蔗从种植到收割、宿根甘蔗从上茬甘蔗收割到宿根甘蔗收割的生长期长短是影响单产的重要因素之一。如将收割期延迟谋求当年最大收益,有可能造成下年产量降低;其次,未完全成熟的春植甘蔗收获过早也将给整体带来损失。因

此,寻求最佳的收割顺序、收割期对提高甘蔗产量和收益及劳动效率具有重要意义。

在不同栽培法对甘蔗单产和含糖量影响研究中,对不同栽培法的种植期及收割期与甘蔗产量和含糖量的关系进行了分析,提出新植蔗及宿根蔗产量和含糖量的预测数学模型<sup>[1]</sup>。本研究根据以上结果,对甘蔗栽培最佳收割期、收割顺序、不同栽培方法的最佳栽培面积等进行了模拟计算,应用遗传算法(Genetic Algorithm, GA)解决甘蔗生产中收割顺序最

孙丽娅(1960~),女,博士,副教授,主要从事系统工程、环境保护和生物能利用的研究。E-mail: sunliya@jssve.edu.cn  
收稿日期: 2009-05-08 接受日期: 2009-07-10

优化的问题，以期为我国甘蔗产业的生产管理提供依据。

### 1 研究方法

决定收割顺序时要考虑当年及下年的栽培方法以及收割田至制糖厂距离、田块间距离、劳动力状况、机械利用等诸多因素的影响<sup>[2,3]</sup>。把所有这些因素同时作为变量求最佳收割顺序是非常困难的，因此，分别求出各个因素下最佳收割顺序，然后再综合分析。

在不同栽培法种植期、收割期与单产、含糖量关系以及新植、宿根甘蔗产量、含糖量数学模型的基础上，为确定最佳收割顺序、收割开始期、不同栽培方法面积，利用 GA 进行了解析。为了使 GA 计算能够较快收敛，对问题进行了格式化、基本设定；对栽培方法推移、栽培方法构成项量、行列式的特性进行了定性分析，开发了适用本研究的自己交叉具体解析法。

### 2 结果与分析

#### 2.1 收割顺序最优化问题及解析法

##### 2.1.1 收割顺序设定

对数年甘蔗收割顺序进行分析，其推移状态如图 1 所示。由图 1 可知，收获顺序与某一个年份无关，可用一个定数代表。为便于解析，首先将问题简单化。收割期将每天收割多种栽培方法的甘蔗设定为只收割 1 种栽培方法的甘蔗。总收割天数  $n$ ，沿收割日期用字符格式编码来表示收割顺序(图 2)。收割顺序最优化最终归结为怎样组合才能得到最大目的

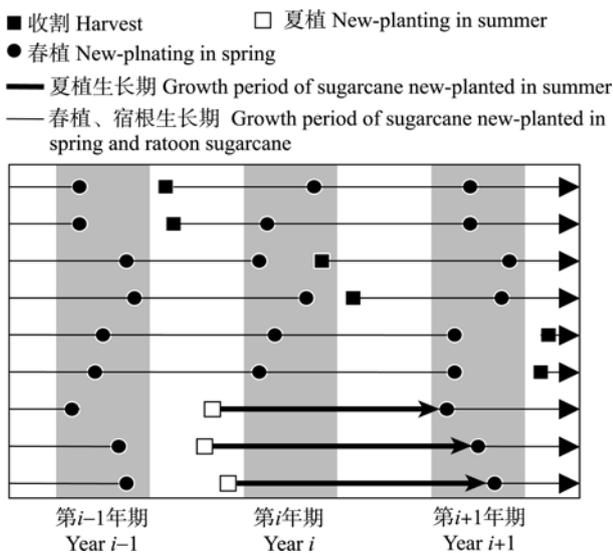


图1 从多年种植及收割考虑收割顺序最佳化  
Fig. 1 Optimization of harvesting sequence on the base of planting and harvest of several years

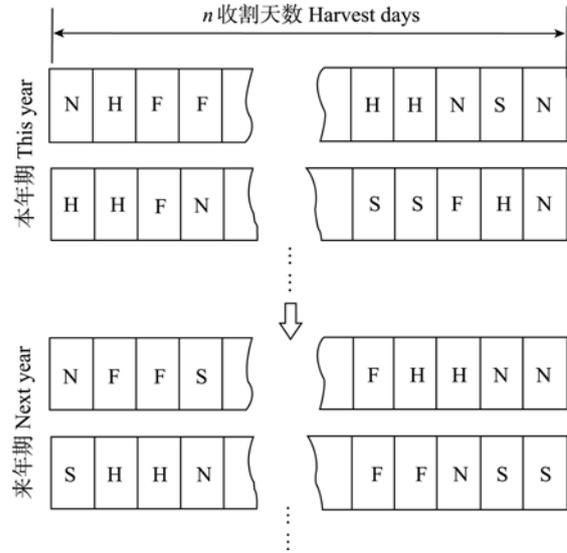


图2 问题模式化  
Fig. 2 Modeling of the problem

N: 夏植 New-planting in summer; H: 春植 New-planting in spring; F: 宿根 1 次 Ratoon once; S: 宿根 2 次 Ratoon twice. 下同 The same below.

函数值(以下称适应度)，也就是组合最优化的问题。

从过去的的数据得知<sup>[4]</sup>，甘蔗栽培方法中，新植的春植、夏植及宿根 1 次、宿根 2 次占甘蔗总种植面积 90%以上，因此本研究对象限定为这 4 种栽培方法，分别用 H、N、F、S 表示；收割天数分别用  $n_H$ 、 $n_N$ 、 $n_F$ 、 $n_S$  表示，总收割天数用  $n$  表示。4 种栽培方法收割天数中总有 3 个是独立的。因此，收割顺序可排列的组合数  $L$  如下：

$$L = \frac{n!}{n_H! n_N! n_F! n_S!} \quad (1)$$

由式(1)得知，可组合数巨大，一般很难求出最佳解。

##### 2.1.2 利用 GA 解收割顺序最优化

GA 是一种有效的解决最优化问题的方法<sup>[5,6]</sup>。利用 GA 算法解收割顺序最优化问题，首先随机挑选一些按收割顺序配置的字符串，将其视为第一代个体，并计算每个解的目标函数值。利用选择机制从个体中随机挑选配置的字符串作为繁殖过程前的样本。选择应保证适应度较高的能够保留较多的样本；而适应度较低解的个体则被淘汰。其次进行交叉操作，这个操作形成下一世代的个体群；交叉是将所对应的位置字符进行替换，在一般 GA 算法中，交叉是在不同个体间进行。由于此方法不适合本研究，故采用交叉在同一个体内进行。在世代交替中，将适应度高的个体样本进行复制，向具有更佳解的个体群推移。由此反复进行适应度评价、个体排列、选别、交叉等操作。在反复计算中，有可能会出现

陷入一个局限的极大值里(局部解)。因此, 时常发生突然变异, 脱出局部解。具有最大优化度特征的个体向下世代推移。趋于一定值时, 再使其发生突然变异, 突然变异又相当于生成了新的个体, 反复多次, 当解完全安定时, 终结计算。

## 2.2 GA 算法求收割程序最优化

### 2.2.1 栽培方法的推移

栽培方法推移的表示法 GA 计算前应先明确个体推移状态, 本年期第  $i$  日收割的栽培方法推移到下期第  $j$  日的另一种栽培方法, 这种持续下去的个体之间有相对应关系; 其次确定下年期新植春植甘蔗的种植日、宿根开始日。以  $A_N$ 、 $A_H$ 、 $A_F$ 、 $A_S$  表示不同栽培法的收获面积, 下年期这些值依栽培方法而延续、变化。不同年份收割面积用  $A_N(1)$ 、 $A_N(2)$ 、 $A_N(3)$  表示, 括号内的数值为年份。夏植甘蔗种植后的第 2 个收割期收割, 即本期收割的夏植甘蔗是去年夏天种植。为易区别, 前者用  $N$ 、后者以  $M$  表示, 如今年不收割的夏植甘蔗面积为  $A_M(1)$ , 本期其他栽培法收割后今年夏植的甘蔗面积为  $A_M(2)$ 。不同栽培法的栽培面积一般以下面的关系式表示:

$$A_N(1) + A_H(1) + A_F(1) + A_S(1) + A_M(1) \neq A_N(2) + A_H(2) + A_F(2) + A_S(2) + A_M(2) \quad (2)$$

但就全地区而言, 甘蔗种植面积一定时, 下式成立:

$$A_N(1) + A_H(1) + A_F(1) + A_S(1) + A_M(1) = A_N(2) + A_H(2) + A_F(2) + A_S(2) + A_M(2) \quad (3)$$

从本期栽培方法推移到下期栽培方法, 以下的行列式成立:

$$\begin{pmatrix} A_N(2) \\ A_H(2) \\ A_F(2) \\ A_S(2) \\ A_M(2) \end{pmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} & t_{15} \\ t_{21} & t_{22} & t_{23} & t_{24} & t_{25} \\ t_{31} & t_{32} & t_{33} & t_{34} & t_{35} \\ t_{41} & t_{42} & t_{43} & t_{44} & t_{45} \\ t_{51} & t_{52} & t_{53} & t_{54} & t_{55} \end{bmatrix} \begin{pmatrix} A_N(1) \\ A_H(1) \\ A_F(1) \\ A_S(1) \\ A_M(1) \end{pmatrix} \quad (4)$$

$$A(2) = TA(1) \quad (5)$$

式中,  $A$  为栽培方法构成的矩阵,  $A = A(A_N, A_H, A_F, A_S, A_M)$ ;  $T$  为栽培方法推移的行列式, 其数列要素  $t_{ij}$  是从栽培方法  $j$  推移到栽培方法  $i$  的比例,  $i$ 、 $j$  为栽培方法, 其中栽培方法  $N$ 、 $H$ 、 $F$ 、 $S$ 、 $M$  分别用 1、2、3、4、5 表示。即使宿根次数增加, 只需相应地将行列式次元增加, 表达式仍然成立。从栽培方法完全没有推移时  $t_{ij}=0$ , 全推移时  $t_{ij}=1$ 。除去  $t_{ij}<0$  及  $t_{ij}>1$ , 栽培方法推移行列式的元素  $t_{ij}$  应满足以下条件:

$$0 \leq t_{ij} \leq 1 \quad (6)$$

式(5)和式(6)表示本年期向下年推移的栽培方法, 对数年推移的状态, 可用下式表示:

$$A(n) = T(n-1) \quad T(2)T(1)A(1) \quad (7)$$

$T(i)$  为从  $i$  年期向第  $i+1$  年期推移的行列式, 对一个地区来说, 当每年甘蔗栽培面积一定时下式成立:

$$A(n) = T^{n-1}A(1) \quad (8)$$

栽培方法推移行列式的特性 栽培方法向下年期推移时具有特定的方向性, 如宿根 2 宿根 1(表示推移方向)( $S \rightarrow F$ )、夏植 宿根 2( $N \rightarrow S$ )、春植宿根 2( $H \rightarrow S$ )的推移不符合逻辑。其次今年不收割夏植甘蔗, 只能明年收割夏植即  $M \rightarrow N$ 。由栽培方法特性决定了行列式的多个要素, 推移式(4)可改为:

$$\begin{pmatrix} A_N(2) \\ A_H(2) \\ A_F(2) \\ A_S(2) \\ A_M(2) \end{pmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ t_{21} & t_{22} & t_{23} & t_{24} & 0 \\ t_{31} & t_{32} & 0 & 0 & 0 \\ 0 & 0 & t_{43} & 0 & 0 \\ t_{51} & t_{52} & t_{53} & t_{54} & 0 \end{bmatrix} \begin{pmatrix} A_N(1) \\ A_H(1) \\ A_F(1) \\ A_S(1) \\ A_M(1) \end{pmatrix} \quad (9)$$

其次由式(4), 本期构成数列中的要素向下年期数列要素推移时,  $i$  最大为 5( $N, H, F, S, M$ ), 所有的列  $j$  应使下式成立:

$$\sum_i t_{ij} = 1 \quad (10)$$

由式(9)、式(10)得出未知数比方程式多, 因此还需再加条件限定决定推移行列式。从经济性上, 夏植 新夏植( $N \rightarrow M$ )、春植 新春植( $H \rightarrow H$ )、春植 新夏植( $H \rightarrow M$ )的栽培方法不合理, 应被排除。从经济性和甘蔗栽培方法的最优化等方面考虑<sup>[7]</sup>, 对栽培方法推移作更进一步限定。

$$\begin{cases} N \rightarrow F, H \\ H \rightarrow F \\ F \rightarrow S, H, M \\ S \rightarrow H, M \end{cases} \quad (11)$$

栽培方法矩阵构成 栽培方法的构成矩阵  $A$  的各元素值由经济性、劳动力、品种特性等要素决定。虽然其构成每年都有变化, 但变化量较小, 可以常量推移。

$$A(1) = A(2) \quad (12)$$

夏植甘蔗栽培面积每年基本保持一定, 因此可认为  $A_M = A_N$ 。其次, 在整个收获期间, 将每天收割面积作为单位面积, 此处设定为 10 hm<sup>2</sup>, 总收割天数为  $n$ , 收割总面积为 10n, 下式关系成立:

$$10n = A_N(1) + A_H(1) + A_F(1) + A_S(1) = A_N(2) + A_H(2) + A_F(2) + A_S(2) \quad (13)$$

各栽培方法的收割天数  $n_N$ 、 $n_H$ 、 $n_F$ 、 $n_S$  分别成为  $A_N/10$ 、 $A_H/10$ 、 $A_F/10$ 、 $A_S/10$ 。由此, 各栽培方法的收割天数相当于栽培方法构成行列式的要素。与之相结合, 栽培方法构成行列式的要素作如下改变, 用  $a_N$ 、 $a_H$ 、 $a_F$ 、 $a_S$  及  $a_M=(a_N)$  表示。

个体的产生 给产生随机数区间赋予不同系数选择栽培方法, 随机产生按收割天数排列的个体。整个收割期各个栽培方法的收割天数与设定值一致。此外, 计算机识别数字比识别文字更方便, 因此, 计算时, 将栽培方法  $N$ 、 $H$ 、 $F$ 、 $S$  分别用 1、2、3、4 来替换, 生成的个体为数字列。每次计算发生的个体数  $m$  由计算机随机产生。

栽培方法推移的决定 在具体表述栽培方法推移时, 必须首先决定推移行列式中的未知元素  $t_{ij}$ , 式(9)可改写为下式:

$$\begin{pmatrix} a_N \\ a_H \\ a_F \\ a_S \\ a_M \end{pmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ t_{21} & 0 & t_{23} & t_{24} & 0 \\ t_{31} & 1 & 0 & 0 & 0 \\ 0 & 0 & t_{43} & 0 & 0 \\ 0 & 0 & t_{53} & t_{54} & 0 \end{bmatrix} \begin{pmatrix} a_N \\ a_H \\ a_F \\ a_S \\ a_M \end{pmatrix} \quad (14)$$

参考(11)式得到以下联立方程:

$$\begin{cases} a_N = a_M \\ a_N t_{21} + a_F t_{23} + a_S t_{24} = a_H \\ a_N t_{31} + a_H = a_F \\ a_F t_{43} = a_S \\ a_F t_{53} + a_S t_{54} = a_N \\ t_{21} + t_{31} = 1 \\ t_{23} + t_{43} + t_{53} = 1 \\ t_{24} + t_{54} = 1 \end{cases} \quad (15)$$

解后得到以下关系:

$$t_{43} = a_S / a_F \quad (16)$$

$$t_{31} = (a_F - a_H) / a_N \quad (17)$$

$$t_{21} = 1 - t_{31} \quad (18)$$

$$a_F t_{23} + a_S t_{24} = a_F - a_N \quad (19)$$

$$t_{54} = 1 - t_{24} \quad (20)$$

$$t_{53} = 1 - t_{23} - t_{43} \quad (21)$$

式(19)中含 2 个未知元素, 是宿根 新植春植( $F$ 、 $H$ 、 $S$ 、 $H$ )的推移式, 未知元素  $t_{23}$  和  $t_{24}$  若不能决定,  $t_{53}$  和  $t_{54}$  也不能决定。由式(7)和  $t_{23}$ 、 $t_{24}$  的关系(图 3),  $a_F$  和  $a_N$  为非负的整数, 式(19)中  $a_F t_{23}$  和  $a_S t_{24}$  分别在等于  $a_F - a_N$  非负的整数中选取。

推移行列式各元素一旦决定, 可决定本期的某栽培方法向下年期推移的栽培方法面积或收割天数。

个体的定常性 从整个地区甘蔗栽培出发, 不同年份的栽培方法构成项量, 构成行列式相对是一个定常的量。因此, 可以认为每一年度任意个体排列都是相同的。

$$\text{本期: } K(1) = (413112223244\dots) \quad (22)$$

$$\text{下年期: } K(2) = (413112223244\dots) \quad (23)$$

$$K(1) = K(2) \quad (24)$$

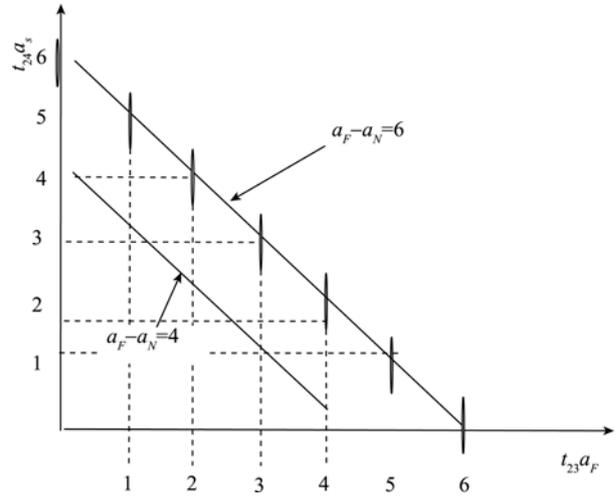


图3  $t_{24}$ 与 $t_{23}$ 的关系  
Fig. 3 Relationship between  $t_{24}$  and  $t_{23}$

栽培方法在推移时的优先顺序 计算中同一栽培方法推移中, 有推移至春植和夏植时, 设定移至春植的早期收割, 夏植的收割期后移。如  $N$ 、 $F$ 、 $H$  中, 早收获的使其优先向春植推移。把个体的世代间所对应的项量用  $V = (v_1 \ v_2 \ v_3 \ \wedge)$  表示。  $v_i$  为第  $i$  日的栽培方法向下年第  $j$  日期推移(图 4)。向宿根( $F$ 、 $S$ )推移时宿根的开始日、向春植  $H$  推移时春植的种植日期可以指定。

□ 夏植 New-planting in summer ● 收割 Harvest  
■ 春植 New-planting in spring

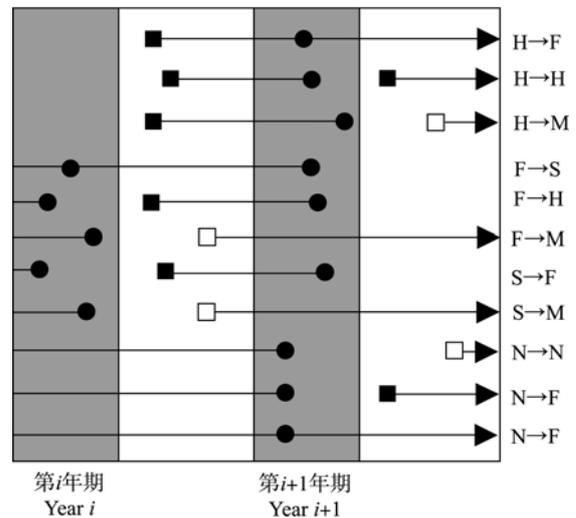


图4 栽培方法推移  
Fig. 4 Transitions of planting modes

### 2.2.2 应用 GA 算法的说明

适应度 适应度可以是总产量、产糖量(单产 × 甘蔗含糖量)或是粗收益的总和。总产量为每天收割量的总合, 产糖量与每天收割量和收割甘蔗的含糖量有关, 呈非线性关系; 粗收益可分为蔗农总粗收益( 单价 × 甘蔗总重量)和制糖厂总粗收益( 单价

× 甘蔗总重量 × 含糖量)。设收割第  $i$  天的单产为  $y_i$ , 含糖量为  $s_i$ , 下式可计算当天的产糖量(粗收益):

$$p_i = y_i f(s_i) \quad (25)$$

适应度  $Z$  为整个收获期间产糖量的总和:

$$Z = \sum_n p_i \quad (26)$$

以多年期( $j$  年期)产糖量进行最优化度评价时:

$$Z = Z(1) + Z(2) + Z(3) + \dots + Z(j) \quad (27)$$

设定个体  $K$  在世代间是定常量[式(17)~式(25)], 用式(27)可以满足评价适应度, 本研究将两年期的合计作为适应度, 仅考虑单产不考虑糖度的适应度计算方法如下:

$$Z = \sum_n y_i f_{\text{const}} \quad (28)$$

$f_{\text{const}}$  为不计含糖量的 1 t 甘蔗的单价。

**个体的选择** 所有个体适应度算出后, 按高值向低值进行排列, 为使适应度高值的易被选择留下, 对其中的复数排列进行复制, 并替换到部分个体中。由此操作, 将适应度高的个体遗传子容易遗传给下一代。

**自己交叉** 与通常 GA 算法不同, 本研究的交叉如式(29)所示,  $i$  与  $j$  在同一个体内进行交替的操作, 称为自己交叉。

$$\begin{matrix} k(k_1 \dots k_i \dots k_j \dots) \\ \Downarrow \\ k(k_1 \dots k_j \dots k_i \dots) \end{matrix} \quad (29)$$

通常 GA 算法中, 交叉在不同个体间互相交替进行。这样使栽培方法的构成比例发生变化, 破坏计算的前提条件。除了最大适应度的个体不进行交叉外, 按顺序设定不同的交叉次数。

过渡状态和定常状态 进行模拟计算的前提条件是收割期、个体、栽培方法构成及栽培方法推移恒定。但 GA 最初的计算条件不明确, 只能赋以特定值使计算开始, 从第 2 年期才能进入安定状态(图 5)称定常状态。因此, 第 1 年期作为过渡状态, 进行评价时将两期的合并作为一个适应度。

### 3 结论

本文探讨了收割顺序及收割时期最优化问题, 从栽培方法特性的角度出发归结为组合最优化问题

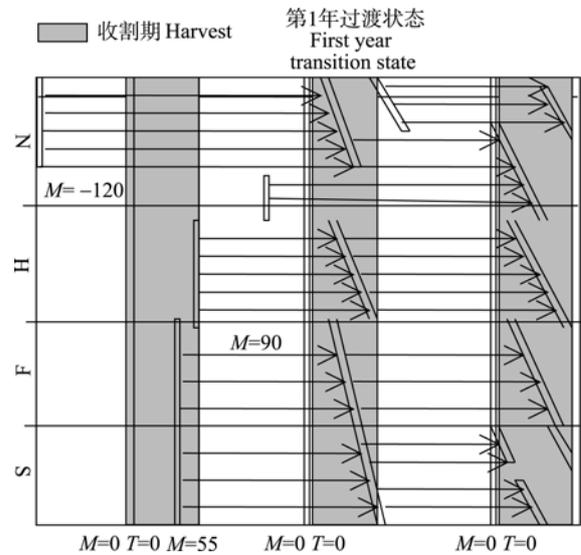


图5 过渡状态到安定状态  
Fig. 5 Shift from transition state to steady state

以及遗传算法(GA)的适用性。在对几个基本设定及对栽培方法推移、栽培方法构成项量、行列式的特性进行了定性分析基础上, 开发了适用本研究的自己交叉, 设第 1 收割期为过渡状态、第 2 收割期为定常状态的目标函数值的和为适应度, 建立遗传算法(GA)的具体解析法。解析法的建立, 结合不同栽培方法下甘蔗单产及含糖量相应的数学模型, 为确定甘蔗栽培最佳收割期、收割顺序、不同栽培方法的最佳栽培面积等模拟计算提供了解析方法。

### 参考文献

- [1] 孙丽娅, 上野正实, 永田雅辉. 不同栽培法对甘蔗单产和含糖量影响的研究[J]. 中国生态农业学报, 2009, 18(5): 992-996
- [2] 佐木隆显, 古谷立美, 南善行. 遗传算法在饲料多目的配合设计中[J]. 农业设施, 1999, 30 (2): 163-171
- [3] 大土井克明, 笈田昭, 山崎焯, 等. 有关农作业计划最优化的研究[J]. 农业机械学会志, 1999, 61 (1): 91-97
- [4] 冲绳糖业振兴会. 甘蔗品质测定结果汇集报告书[R]. 冲绳: 冲绳糖业振兴会, 1995-2007
- [5] 古田均, 山本博之. 遗传算法在构造工学的应用[M]. 东京: 森北出版社, 1997: 5-35
- [6] 周明, 孙树栋. 遗传算法原理及应用[M]. 北京: 国防工业出版社, 2002: 12-30
- [7] 金城铁男. 甘蔗栽培方法的最优化[R]//甘蔗关系试验成绩概要书. 冲绳: 冲绳农林水产部, 1997: 87-92